A CORRELATIONAL APPROACH TO PREDICTING OPERATOR STATUS

Clark A. Shingledecker
NTI, Incorporated
Dayton, Ohio

ABSTRACT

This paper discusses a research approach for identifying and validating candidate physiological and behavioral parameters which can be used to predict the performance capabilities of aircrew and other system operators. In this methodology, concurrent and advance correlations are computed between predictor values and criterion performance measures. Continuous performance and sleep loss are used as stressors to promote performance variation. Preliminary data are presented which suggest dependence of prediction capability on the resource allocation policy of the operator.

INTRODUCTION

Modern advances in engineering and electronics technology continue to be responsible for a phenomenal increase in the potential effectiveness of military and commercial aircraft systems. However, the enhanced speed, operating range, maneuverability, remote sensing, and weapons capabilities made possible by these technologies are also producing significant changes in the role and importance of critical flight crew members, and in the performance requirements that are imposed upon them. As a consequence, serious consideration must be given to methods and approaches which can be used to insure optimal human performance in future airborne operations.

Several factors contribute to a growing concern over the maintenance of aircrew performance. The use of increasingly sophisticated flight computers has relieved the aircrew of many labor-intensive duties, and shifted their task to one of monitoring and supervising a complex and highly flexible system. Such automation often leads to a reduction in crew size and creates a situation in which increasingly critical responsibilities are assigned to individual operators whose performance can easily become the single most important determinant of the outcome of a major battle or of the safety of hundreds of passengers.

The problem of reduced crew redundancy is compounded by a concommitant increase in mental workload. The cockpits and flight decks of contemporary aircraft are capable of providing pilots with vast amounts of data that must be processed in a timely and accurate manner if system performance is to be maintained. In many cases, the resulting perceptual and cognitive task demands can approach, and even exceed, the inherently limited information processing capacities of even the most experienced personnel.

Traditionally, human factors specialists have approached the problem of supporting pilot performance through the design of crew station interfaces to minimize information overload, and through the development of improved training technologies. While these interventions have been successful, it is unlikely that they will continue to be sufficient by themselves to insure optimal system performance in an environment where pilot task demands are increasing, and pilot performance capabilities can be degraded by a variety of physical and psychological stressors. Included among the obvious threats to aircrew performance capacities are fatigue and sleep loss in extended operations, use of prescribed or illegal drugs, and in combat aircrews, exposure to chemical, biological and nuclear threats.

Taken together, the rising criticality of the performance exhibited by key crew members, growing task demands and the incapacitating potential of operational stressors suggest that specific, interactive subsystems may be needed to guard against catastrophic failures due to human error.

One technically feasible approach that has been suggested for preventing human errors would involve monitoring the performance capabilities of the human operator. At the simplest level, such biocybernetic intervention would permit the evaluation of performance capability prior to a flight in order to select those personnel who exhibit an optimal capacity to meet mission objectives. In a more advanced application, performance capabilities could be monitored on a moment-to-moment basis during a mission. Thus, impending operator performance decrements could be detected automatically, and the information used to alert the pilot, inform command personnel or even initiate computer control of the system.

The general computer hardware, software, and sensing technology is currently available to implement biocybernetic systems capable of monitoring the performance capability of human operators. However, little is presently known about the indices of human function that could be used to accurately and reliably measure and predict performance capabilities in a non-intrusive fashion. The purpose of this paper is to present a methodological approach with preliminary data aimed at identifying behavioral and electrophysiological predictors of impending performance failure.

RESEARCH METHOD

The methodology developed for this exploratory research represents a departure from classical research techniques which are employed to investigate measures of performance capability. In such traditional studies, the goal is to assess a measure's capability to reflect the presumed impact of an intervening hypothetical construct (e.g., fatigue, chemical intoxication, boredom, disease) on the human operator. Thus, these studies

attempt to show that when an independent variable such as sleep loss or time-on-task is varied, the measure under examination behaves in a manner which is hypothesized to be functionally equivalent to a concommitant change in the intervening variable (e.g., a monotonic increase in reaction time with increasing fatigue).

While such experimental approaches are acceptable in research designed to investigate specific pscychological phenomena, they are neither warranted nor appropriate when the research goal is to identify measures which predict performance change. The purpose of the methodology demonstrated in the present study is to specify metrics that predict performance variation. This purpose dictates a more operational approach where, rather than testing a hypothesis about causal factors linking an intervening variable and performance, a relationship is sought between a predictor metric and a criterion performance index.

In the present methodology, candidate performance predictor metrics are correlated with simultaneous and temporally succeeding measures of performance on a simulated systems operation task. Within this approach, predictor measures which correlate highly with performance on the criterion or primary task of interest can be considered reliable indicators of operator performance decrement.

While human performance naturally varies within a restricted range under normal conditions, the degree of variation observable over a typical experimental session is likely to be highly constrained. Thus, in the present methodology, performance variability is induced by exposing subjects to the combined stressors of sleep loss and continuous performance. It should be noted that the intent of imposing these stressors is not to produce some predicted pattern of decrement due to fatigue or diurnal cycles of performance efficiency. Instead, the technique is simply designed to capitalize on the performance variation likely to be produced by these conditions in order to examine a broad range of within-subject performance variability.

In summary, the object of the methodology is to provide a standardized approach to evaluating candidate measures which will predict reductions in performance capability. The approach is essentially correlational and is designed to provide quantitative estimates of the capacity of physiological, behavioral or subjective metrics to predict the variability of human performance on a task of interest.

A limited experimental implementation of the methodolgy has been completed in which two subjects performed a complex time sharing task continuously for eight hours following twelve preceding hours of sleep deprivation. This task was designed to simulate a generic systems operation activity (e.g., combat aircraft operation) and contained two primary components which

were performed simultaneously with equal priority. The first of these components was a manual control task.

The control task was a single axis (vertical), unstable compensatory tracking task similar to that described by Shingledecker (ref. 1). The task required subjects to view a cursor on a monochrome video monitor, and to keep the cursor centered over a fixed target by turning a control knob.

The second component of the simulated operational task was a visual monitoring task. The monitoring task is somewhat similar to that devised by Alluisi (ref. 2) and requires subjects to view four computer generated vertical displays that are similar to tape instruments. The scale on each display consists of six hash marks, and the center of the scale is indicated by a small circle. Under nonsignal conditions, the pointers located just to the left of the scale markings on each dial move from one position to another in a random fashion. The pointer movements on each dial are totally independent of the other dials, and occur at an update rate of 5 moves/sec. At unpredictable time intervals, the pointer on one of the four dials becomes biased to either the top half or the bottom half of the scale. This signifies a signal condition to which the subject is instructed to respond by pressing the appropriate key on a four-button keypad. Signals occurred at a frequency of 4 to 5 each minute.

To perform the combined tasks, the subject sat at a work station containing two video monitors. The tracking task was displayed on a screen which was located directly in front of the subject. The monitoring task was displayed on a monitor centered above the tracking monitor and tilted approximately 20 degrees toward the subject. Viewing distance for both monitors was approximately 60cm. The tracking task was controlled by rotating a knob in the horizontal plane with the dominant hand. The monitoring task responses were recorded from four push buttons controlled by the non-dominant hand.

Five candidate predictor measures were selected to match the information processing demands of the system operation task. In order to assess general activation level factors, four frequency bands of the EEG spectrum were selected for power spectrum analysis. In addition, as general measures of alertness, eyeblink closure duration and subjective fatigue metrics were employed.

A primary aspect of the simulated systems operation task was a display monitoring activity. In order to assess such perceptual demands, the visual memory search task was selected (Sternberg, ref. 3). Finally, in order to assess the response output capabilities of the operator associated with the high manual control demands of the vehicle operation task, the Interval Production Task (IPT) was used (Michon, ref. 4).

RESULTS

Data were collected on the criterion systems operation task and on the physiological metrics in five minute intervals. The interpolated behavioral measures were collected during a break period preceding each 50 minute performance period. Advance correlations between the predictor measures and criterion performance were computed for a variety of temporal relationships. However, to permit comparisons across the behavioral and physiological measures, only advance predictor correlations for the eight performance periods are discussed here. In this case, predictive relationships were assessed by correlating mean tracking and monitoring scores for each hour with the physiological metrics obtained in the preceding hour, or with the behavioral data collected during the preceding break period.

These correlations are shown in Table 1. Although the results are based on only two subjects, a number of tentative observations can be made from these data regarding the relative predictive capacity of the candidate parameters.

A strong relationship was obtained between performance and the proportion of total EEG power in each of four measured frequency bands. As shown in Table 1., both tracking error and monitoring signal misses were associated with power in each band. The pattern of correlation across the four bands is a general shift in power, such that poorer criterion performance occurred when the relative power in the low frequency band (delta, 1-3 hz) increased and relative power in higher frequency bands decreased (4-30 hz).

Similarly positive predictive relationships were obtained for the measures of eyeblink behavior. Increases in tracking error as well as poorer signal detection were predicted by larger amplitude blinks, higher blink rates, longer descent times for the eyelid, and longer closure durations.

A more variable set of relationships was obtained between the interpolated behavioral task measures and criterion performance. In general, criterion task decrements were associated with a decrease in duration of the interval between finger taps on the IPT task and an increase in the variability of intertap intervals. Longer Sternberg memory search task reaction times were also predictive of poorer criterion performance.

Although the results summarized above are generally descriptive of the average correlations between the predictor measures and criterion performance, inspection of Table 1 reveals marked individual differences between the two pilot subjects. Specifically, for Subject 1, correlation coefficients were consistently larger for the monitoring performance index

than the tracking. In contrast, the predictor measures were more strongly associated with tracking error than monitoring misses for Subject 2.

A potential explanation for this finding is apparent in an inspection of the hourly mean performance scores that were recorded on the two elements of the simulated systems operation task. Over the eight hour testing period, Subject 1 displayed no more than a 22% variation in tracking error. In contrast, monitoring performance varied as much as 60% and declined consistently across the testing sessions. The opposite pattern of performance was apparent for Subject 2 who displayed a greater decrement in tracking performance. Since the time sharing nature of the criterion task allowed the subjects to freely allocate their attentional resources to the tracking and monitoring components, these data suggest that the subjects devoted the bulk of their diminishing capacities to different components of the criterion task.

Such an explanation is congruent with the correlational findings for the physiological and behavioral predictors. Apparently, for these metrics predictive power may be dependent on the resource allocation policy adopted by the performer. Thus, in the case of the pilot subjects, performance on the interpolated behavioral tasks anticipated the component of criterion task performance that received the least effort expenditure. In support of this interpretation, subjective fatigue ratings for Subject 1 were positively related to monitoring missed detections (r=.92), but unrelated to tracking errors (r=-.02). Likewise, for Subject 2, fatigue ratings were strongly associated with tracking error (r=.92), but were not significantly correlated with monitoring misses (r=.26).

CONCLUSIONS

The results outlined above suggest that the methodological approach described in this paper can be used to identify and select reliable indicators of impending performance degradation in aircrews and in the operators of other critical systems. In order to develop practical technologies for monitoring human performance capabilities, a focused effort will be required in which these techniques are exercised to specify useful parameters, to validate their predictive capabilities for operational situations, and to embody them in field-usable hardware.

The work reported here suggests that no single index of human function is likely to provide global performance prediction in all task environments. Thus, accurate anticipation of performance degradation will probably be achieved only by a family of technologies from which appropriate measures will be selected to match operational environments. At a minimum, such matching will be based on three groups of factors.

350

As suggested by multi-factor models of human performance, a primary consideration will be the information processing resource structure of the operator's task. Measures which assess the integrity of perceptual, central and response processes as well as activation level will have to be selectively applied to tasks and environments which make differential demands on these resources. In addition, as the present results indicate, task priorities will have to be assessed in order to determine the specific aspects of performance that will be predicted by monitoring parameters.

A second group of matching factors is the temporal prediction requirement of the operational scenario. The complete results of the preliminary study indicated that different metrics varied in terms of the time period for which significant predictions were obtained. Thus, it will be necessary to employ these measurement methods in a selective manner to correspond with requirements for long term predictions (e.g. how likely is it that pilot "A"'s performance will be degraded in the next five hours?) and for short term, continuous prediction (e.g., is it probable that pilot "B" will commit a catastrophic error in the next few minutes?).

Finally, selection of prediction measures will also be determined by the limits and practicalities of the operational environment. For example, the potential intrusiveness of some measures may prevent their use during high demand, continuous performance missions. However, these measures may be preferable in situations where periodic, interpolated testing is possible. Other practical selection factors might include the size and weight of the monitoring equipment, and the operator's acceptance of any necessary monitoring sensors.

REFERENCES

1. Shingledecker, C. A. A Task Battery for Applied Human Performance Assessment Research. Air Force Aerospace Medical Research Laboratory, AFAMRL-TR-84-071, 1984.

2. Alluisi, Earl A. Methodology in the use of synthetic tasks to assess complex performance. Human Factors, 1967, 9 (4), 375-384.

3. Sternberg, S. Memory Scanning: Mental processes revealed by reaction time experiments. American Scientist, 1969, 57, 421-457.

4. Michon, J. A., Tapping regularity as a measure of perceptual motor load. Ergonomics, 1966, 9 (5), 401-412.

TABLE 1.

Performance Prediction Correlations

EEG Proportional Power

		Delta	Theta	Alpha	Beta
Tracking Error	S1	.14	.29	-.17	-.21
	S2	.93	-.88	-.94	-.93
Missed Signals	S1	.95	-.64	-.86	-.97
	S2	.17	-.07	-.26	-.29

EOG Eyeblink Parameters

		Interval	Amplitude	Duration	Descent
Tracking Error	S1	-.14	.22	.37	0
	S2	-.76	.10	.99	.94
Missed Signals	S1	-.84	.62	.80	.81
	S2	-.22	-.02	.04	-.20

Interpolated Behavioral Tests

		IPT Duration	IPT Variability	Sternberg RT
Tracking Error	S1	-.28	.22	.01
	S2	-.66	.69	.35
Missed Signals	S1	-.76	.53	.55
	S2	-.30	-.05	-.07